Solutions to the Problems in
Statistics for Engineers and Scientists: Fifth Edition
by William Navidi

John Weatherwax

To my family.

Introduction

This is a solution manual to a few of the problems in the excellent textbook:

Statistics for Engineers and Scientists: Fifth Edition
by William Navidi

I'm currently working on finishing more of the problems in this book. In the meantime I'm publishing my partial results for any student who does not want to wait for the full book to be finished.

One of the benefits of this manual is that I heavily use the R statistical language to perform any of the needed numerical computations (rather than do them "by-hand"). Thus if you work though this manual you will be learning the R language at the same time as you learn statistics. The R programming language is one of the most desired skills for anyone who hopes to use data/statistics in their future career. The R code can be found at the following location:

http://waxworksmath.com/Authors/N_Z/Navidi/navidi.html

As a final comment, I've worked hard to make these notes as good as I can, but I have no illusions that they are perfect. If you feel that that there is a better way to accomplish or explain an exercise or derivation presented in these notes; or that one or more of the explanations is unclear, incomplete, or misleading, please tell me. If you find an error of any kind – technical, grammatical, typographical, whatever – please tell me that, too. I'll gladly add to the acknowledgments in later printings the name of the first person to bring each problem to my attention.

Chapter 1: Sampling and Descriptive Statistics

Sampling

Exercise 1

Part (a): The population is all possible runs of the chemical process. This is a conceptual population as not all possible runs could be produced.

Part (b): The population is all possible voters. This is a tangible (however large) population as there is a finite number of votes.

Part (c): The population is all people with high cholesterol. This is a tangible population as there is a finite number of people with this condition.

Part (d): The population is all concrete specimens that could be constructed from this new formulation. This is a conceptual population as in theory we could keep creating specimens.

Part (e): The population is all bolts that are manufactured on a certain day. This is a tangible population since in theory he could consider all of the bolts made.

Exercise 2

One would use method iii. Since our population is all students at the university it is the one that has no "conditions" on the choice of the student selected. This would give a more random sample than the other methods as they are selecting more narrow subsets of the population.

Exercise 3

Part (a): False. A random sample tries to do this but it may not succeed.

Part (b): True. This is the definition of a simple random sample.

Exercise 4

Part (a): False. The sample tells us something about the population but not with certainty.

Part (b): True

Exercise 5

Part (a): Not necessarily. In a sample size of 100 even for a process that is "better" than the original it is possible to get 12 defective circuits.

Part (b): Finding 11 defective circuits is some evidence (but perhaps not much) that the new process might be better. It could even be worse.

Part (c): Only finding two defective circuits is stronger evidence that the new process is better than the existing process.

Exercise 6

Part (a): We cannot conclude this. A sample only adds evidence for or against a hypothesis. In this case the hypothesis is that the proportion of defective circuits is 12%.

Part (b): This is true.

Part (c): This is true.

Exercise 7

A good knowledge of how the data is generated will allow one to determine if a sample is a simple random sample.

Exercise 8

Part (a): This is an observational study since the measurements are made on the people selected who came to the health fair.

Part (b): If we seek to show that exercise reduces blood pressure we should select a sample of participants who have high blood pressure and subject some of them to exercise. We would then need to compare average change in blood pressure between the two groups. What we have done here is to just take a sample, observe a characteristic of that sample, and then jump to a cause and effect relationship. The correct procedure is actually discussed in the next exercise.

Exercise 9

Part (a): This is a controlled experiment.

Part (b): Yes. The only difference between the two groups is the presence or absence of exercise. If the participants are sufficiently random any differences in outcome (i.e. blood pressure) would therefore to be due to exercise.

Summary Statistics

Exercise 1

False. Half of them will be below the median. The mean can be a "skewed" statistic.

Exercise 2

No it is not. Consider the data $1, 1, 1, 1, 100$ which has a sample mean of $\bar{X} = 20.8$ while the most frequent value is one.

Exercise 3

No. See the data example in the previous exercise.

Exercise 4

Not if the sample size n is even. For example the median of the data $1, 2, 3, 4$ is 2.5.

Exercise 5

Any sample with an odd number will have a median that is a number in the sample as the median in that case is the $\frac{n+1}{2}$th element in the sorted list of numbers.

Exercise 6

I think yes. I think it is more likely to be true if the sample has "outliers". The sample $1, 1, 1, 1, 100$ has a mean of 20.8 and a standard deviation of 44.27415.

Exercise 7

Only if the numbers are all constant. Thus the data $1, 1, 1, 1$ has a standard deviation of zero.

Exercise 8

The mean salary will go up by \$50 and the standard deviation will not change.

Exercise 9

This means that each measurement s_i of salary becomes $1.05s_i$. This means that the mean salary will increase by 5% also. The standard deviation will now also be multiplied by 1.05 again an increase of 5%.

Exercise 10

See the `python` file `chap_1_ex_10.py` where the computations for this problem are worked.

Exercise 11

This would be

$$\frac{20(178) + 30(164)}{20 + 30} = 169.6 \,.$$

Exercise 12

See the `python` file `chap_1_ex_12.py` where the computations for this problem are worked. When we run that we get

```
mean  median  first_quartile  third_quartile  standard_deviation
```

```
Method_A  22.74375    23.5        21.500        25.0        2.872390
Method_B  20.70000    20.4        19.975        22.0        1.353514
Method_C  21.01250    21.0        20.775        21.5        0.419325
Method_D  20.80625    20.7        20.150        21.2        0.745179
```

The method with the largest standard deviation is "Method A". This is expected as all measurements are made "by-eye". A smaller standard deviation is better as it means the measurements are more exact.

Exercise 13

Part (a): To apply this transformation we would be dividing the numbers in the column by 2.54. This means that for that column the mean, the median, the quartiles, and the standard deviation would all be divided by 2.54.

Part (b): Not necessarily as the errors would be different each time.

Exercise 14

Part (a): The new mean would be

$$\frac{10(70000) - 100000 + 1000000}{10} = 160000,$$

over double the original mean.

Part (b): The median would not change.

Part (c): Using the formula for s^2 we can write

$$\sum_{i=1}^{n} X_1^2 = s^2(n-1) + n\bar{X}^2$$

Since we know everything on the right-hand-side before the change we have

$$\sum_{i=1}^{n} X_i^2 = 20000^2(9) + 10(70000^2) = 52600000000.$$

If we subtract the old sample (squared) and add in the new sample (squared) we find the new sum of X_i^2 will be

$$52600000000 - 100000^2 + 1000000^2 = 1042600000000.$$

This means that the new standard deviation using these pieces is given by

$$s = 886904.7,$$

much larger than before.

Exercise 15

See the python file `chap_1_ex_15.py` where the computations for this problem are worked. I computed the values of $\frac{k}{n}(n+1)$ and then rounded to the nearest integer to determine the "sample" break points. Other conventions are possible to use in determining the location of the break points. For example we could interpolate between the two nearest points. When we run that python code we get

```
Tertile X[] indices (zeros based):
[ 5 10]
Cut points:
[46, 76]
Qunitiles X[] indices (zeros based):
[ 2  6  9 13]
Cut points:
[23, 49, 74, 89]
```

Exercise 16

Part (a): The outlier in this example i.e. the number 4.91 looks like it was meant to be entered as 49.1 and then it would be in the range of the other numbers.

Part (b): The outlier in this example i.e. the number $110,000 could be due to a very nice/expensive car and not an error in entry.

Chapter 6: Hypothesis Testing

Large-Sample Tests for a Population Mean

Exercise 1

Part (a): We have $\mu_0 = 1$ and compute the z-score

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{1.03 - 1}{0.1/\sqrt{50}} = 2.12132 \,.$$

We then compute the P-value and find $P = 0.01694743$. This is done in the R code

```
z = (1.03 - 1)/(0.1/sqrt(50))
P = pnorm(-z)
```

Part (b): Or in the most extreme of the 1.7% of the distribution.

Exercise 2

Part (a): We have $\mu_0 = 0.6$ and compute the z-score

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{0.595 - 0.6}{0.05/\sqrt{65}} = -0.8062258 \,.$$

We then compute the P-value and find $P = 0.2100563$. This is done in the R code

```
z = (0.595 - 0.6)/(0.05/sqrt(65))
P = pnorm(z)
```

Part (b): Or in the most extreme of the 21% of the distribution.

Exercise 3

Part (a): We have $\mu_0 = 232$ and compute the z-score

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{231.7 - 232}{2.19/\sqrt{66}} = -1.112882 \,.$$

We then compute the P-value and find $P = 0.2657591$. This is done in the R code

```
z = (231.7 - 232)/(2.19/sqrt(66))
P = pnorm(z) + (1 - pnorm(-z))
```

Part (b): Or in the most extreme of the 26.5% of the distribution.

Exercise 4

Part (a): We have $\mu_0 = 2.5$ and compute the z-score

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.6 - 2.5}{0.3/\sqrt{50}} = 2.357023 \,.$$

We then compute the P-value and find $P = 0.009211063$. This is done in the R code

```
z = (2.6 - 2.5)/(0.3/sqrt(50))
P = pnorm(-z)
```

Part (b): Or in the most extreme of the 0.09% of the distribution.

Exercise 5

Part (a): We have $\mu_0 = 5.4$ and compute the z-score

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{4.5 - 5.4}{2.7/\sqrt{80}} = -2.981424 \,.$$

We then compute the P-value and find $P = 0.001434556$. This is done in the R code

```
z = (4.5 - 5.4)/(2.7/sqrt(80))
P = pnorm(z)
```

Part (b): If the mean number of sick days was 5.4 then there is only a $100P = 0.1434556$ percent chance of observing a value for $\bar{X}$ this large by chance (effectively zero). This gives evidence to the idea that the mean number of sick days is *less* than 5.4.

Exercise 6

Part (a): Following the above we first compute the z-score and then the P-value. This is done in the R code

```
z = (15.2 - 15)/(1.8/sqrt(87))
P = 2*pnorm(-z)
```

We find $P = 0.300027$.

Part (b): If the mean diameter was 15 then there is a $100P = 30.0027$ percent chance of observing a value for $\bar{X}$ this large by chance. This indicates we don't have much evidence that $\mu \neq 15$.

Exercise 7

Part (a): Following the above exercises we first compute the z-score and then the P-value. This is done in the R code

```
z = (850 - 900)/(153/sqrt(60))
P = pnorm(z)
```

We find $P = 0.005681031$.

Part (b): If the mean depth was equal to 900 there is only a $100 \times P$ percent chance of observing a value for $\bar{X}$ this small by chance. This indicates that it is very likely that the mean pit depth is less than 900.

Exercise 8

Part (a): Following the above exercises we first compute the z-score and then the P-value. This is done in the R code

```
z = (5.02 - 5)/(0.06/sqrt(100))
P = 2*pnorm(-z)
```

We find $P = 0.0008581207$.

Part (b): If the mean depth was equal to 5 there is only a $100 \times P$ percent chance of observing a value for $\bar{X}$ this far from $\mu_0 = 5$ by chance. This indicates it is very likely that the process is *out* of calibration.

Exercise 9

Part (a): Following the above exercises we first compute the z-score and then the P-value. This is done in the R code

```
z = (8.24 - 10)/(16.33/sqrt(126))
P = pnorm(-abs(z))
```

We find $P = 0.1131788$.

Part (b): There is a $100 \times P = 11.31788\%$ chance of observing a value for $\bar{X}$ this far below $\mu_0 = 10$ by chance. This is not normally enough evidence to make a claim that something is amiss.

Exercise 10

Part (a): Following the above exercises we first compute the z-score and then the P-value. This is done in the R code

```
z = (1356 - 1350)/(70/sqrt(100))
P = 1 - pnorm(z)
```

We find $P = 0.195683$.

Part (b): There is a $100 \times P = 19.5683\%$ chance of observing a value for $\bar{X}$ this extreme by chance. This is not normally enough evidence to make a claim that the new concrete mix has a larger compressive strength.

Exercise 11

The mean of $\bar{X}$ under the null would be five.

Exercise 12

If we repeated this experiment n times we would expect to see a z value this extreme $0.04n$ times. If $n = 100$ this is four times.

Exercise 13

If we repeated this experiment n times we would expect to find a value of $\bar{X}$ between 4 and 6 approximately $(1 - P)n = (1 - 0.22)n = 0.78n$ times. If $n = 100$ this would be 78.

Exercise 14

We compute a z-value and then the P-value using the following R code

```
z = (7.4 - 7)/0.2
P = 2*pnorm(-z)
```

We find $P = 0.04550026$.

Exercise 15

We compute a z-value and then the P-value using the following R code

```
z = (5.0 - 5.2)/0.1
P = 2*pnorm(-abs(z))
```

We find $P = 0.04550026$.

Exercise 16

Part (a): This is a two-tailed test.

Part (b): The null hypothesis is that $H_0 : \mu = 73.5$.

Part (c): The P-value is 0.196.

Part (d): As this is a two-tailed test the P-value for a one-tailed test would be one-half the given two-tailed test P-value or 0.098.

Part (e): We can compute this using the following R code

```
n = 145
```

```
xbar = 73.2461
s = 2.3634
z = qnorm(1-0.01/2)
ci = xbar + c(-1, +1)*z*(s/sqrt(n))
```

This gives the interval $(72.74054, 73.75166)$

Exercise 17

Part (a): This is a one-tailed test.

Part (b): The null hypothesis is that $H_0 : \mu = 45$.

Part (c): The P-value is 0.003.

Part (d): We would need to compute $z = \frac{\bar{X}-46}{\sigma/\sqrt{n}}$. In the R code we find

```
n = 122
xbar = 46.2598
s = 5.0921
z = (xbar - 46)/(s/sqrt(n))
P = pnorm(-abs(z))
```

This gives a P-value of 0.2865348.

Part (e): We can compute this using the following R code (extending the code above)

```
z = qnorm(1-0.02/2)
ci = xbar + c(-1, +1)*z*(s/sqrt(n))
```

This gives the interval $(45.18731, 47.33229)$.

Exercise 18

We can compute these using the following R code

```
s = 2.2136
se_mean = 0.2767
n = ( s/se_mean )^2 ## Using: se_mean = s/sqrt(n)
```

```
z = -1.03
mu0 = 10.5
xbar = se_mean*z + mu0 ## Using: z = (xbar - mu0)/se_mean
P = pnorm(z)
print(sprintf('n= %f; xbar (mean)= %f; P= %f', n, xbar, P))
```

which give

```
[1] "n= 64.000000; xbar (mean)= 10.214999; P= 0.151505"
```

Exercise 19

We can compute these using the following `R` code

```
n = 87
xbar = 4.07114
s = 2.00819
se_mean = s/sqrt(n)
z = (xbar - 3.5)/se_mean
P = 2*pnorm(-abs(z))
print(sprintf('se_mean= %f; z= %f; P= %f', se_mean, z, P))
```

which give

```
[1] "se_mean= 0.215301; z= 2.652757; P= 0.007984"
```

Drawing Conclusions from the Results of Hypothesis Tests

Exercise 1

WWX: Working here.

Chapter 10: Statistical Quality Control

Basic Ideals

Exercise 1

Part (a): This would be a count variable.

Part (b): This would be a continuous variable.

Part (c): This would be a binary variable.

Part (d): This would be a continuous variable.

Exercise 2

Part (a): True.

Part (b): False. A process could be in control but meeting the desired specifications.

Part (c): This is assumed and is thus true.

Part (d): False. Variability within the items sampled should be due to common causes.

Part (e): False. Some variation is expected. What we are seeking to observe is variability that is due to "special causes".

Exercise 3

Part (a): In control.

Part (b): High capability.

Exercise 4

It must still be monitored continually to watch for situations where it might go out of control.

Exercise 5

Part (a): False. It is "hoped" that most of the output will meet specifications but that is more under the definition of process "capability". Process "control" is more about the stability of the process.

Part (b): This might be true but it also might be false.

Part (c): True.

Part (d): True.

Exercise 6

We prefer to sample small units to observe special causes quickly.

Control Charts for Variables

Exercise 1

Part (a): We first need to compute $\bar{R}$. From the numbers given we find $\bar{R} = \frac{143.7}{30} = 4.79$. Then the 3σ control limits for the R chart are given by $D_4\bar{R}$ and $D_3\bar{R}$ with D_4 and D_3 given by tables in the book. We find these to be $(\text{LCL}, \text{UCL}) = (0.0, 10.93078)$.

Part (b): We first need to compute $\bar{s}$. From the numbers given we find $\bar{s} = \frac{62.5}{30} = 2.083333$. Then the 3σ control limits for the S chart are given by $B_4\bar{S}$ and $B_3\bar{S}$ with B_4 and B_3 given by tables in the book. We find these to be $(\text{LCL}, \text{UCL}) = (0.0, 4.720833)$.

Part (c): We compute $\bar{\bar{X}} = \frac{712.5}{30} = 23.75$. Then when using $\bar{R}$ to estimate $\sigma_{\bar{X}}$ the 3σ control limits for $\bar{X}$ are given by

$$(\text{LCL}, \text{UCL}) = (\bar{\bar{X}} - A_2\bar{R}, \bar{\bar{X}} + A_2\bar{R}).$$

Here A_2 depends on n and is given in the book. Computing these we find $(\text{LCL}, \text{UCL}) = (20.25809, 27.24191)$.

Part (d): When using $\bar{s}$ to estimate $\sigma_{\bar{X}}$ the 3σ control limits for $\bar{X}$ are given by

$$(\text{LCL}, \text{UCL}) = (\bar{\bar{X}} - A_3\bar{s}, \bar{\bar{X}} + A_3\bar{s}).$$

Here A_3 depend on n and are given in the book. Computing these we find $(\text{LCL}, \text{UCL}) = (20.35833, 27.14167)$.

The numerical values above were computed with the following R code

```
n = 4

bar_R = 143.7/30
D4 = 2.282
D3 = 0.0
print(c(D3*bar_R, D4*bar_R))

bar_S = 62.5/30
B4 = 2.266
B3 = 0
print(c(B3*bar_S, B4*bar_S))

barbar_X = 712.5/30
A2 = 0.729
print(c(barbar_X - A2*bar_R, barbar_X + A2*bar_R))

A3 = 1.628
print(c(barbar_X - A3*bar_S, barbar_X + A3*bar_S))
```

Exercise 2

We can look at the Western Electric rules one at a time

- No points are outside of the 3σ control limits.

- There are no "two-out-of-three" points above or below the 2σ control limits.

- There are no "four-out-of-five" points above or below the σ control limits.

- There are not eight consecutive points on the same side of the center line.

Based on this we would conclude that the process is in control.

Exercise 3

This problem is worked in the R code chap_10_sect_2_prob_3_N_4.R.

Part (a): Running the above code we see that the variance of the process seems to be in control and no samples need to be deleted.

Part (b): Running the above code we see several $\bar{X}$ values that are outside of the 3σ control limits. I see that samples numbered 8 and 9 have $\bar{X}$ that is lower than the lower confidence bounds for the mean.

Part (c): The sample is deemed out of control on sample 7 where two of three points are beyond the 2σ control limits.

Exercise 4

The computational parts of this problem are implemented in the R code for the previous exercise. When we run that code the standard deviation based 1, 2, and 3σ control limits for $\bar{X}$ are plotted. We again see that the 8th and 9th samples are out of the control limits (in agreement with the observation in the previous exercise). The Western Electric rules would first signal an "out-of-control" warning on sample 7 where again two of three samples are beyond the 2σ (negative) control limits.

Exercise 5

Part (a): If the process mean shifts to 14 the sample means will be generated by

$$\bar{X} \sim \mathcal{N}\left(14, \frac{3^2}{5}\right),$$

while the control limits will be set by the old process mean which would be

$$12 \pm 3\left(\frac{3}{\sqrt{5}}\right).$$

I find these to be $(7.975078, 16.024922)$.

Computing p the probability that a single sample is outside of the confidence intervals gives

```
[1] "p= 0.065616; ARL= 15.240151"
```

Part (b): Let the mean be shifted to upwards m so that

$$\bar{X} \sim \mathcal{N}\left(m, \frac{3^2}{5}\right).$$

Then with $\mathrm{ARL} = 4$ we would then desire to have

$$\frac{1}{\mathrm{ARL}} = P(\bar{X} < 7.975078) + P(\bar{X} > 16.024922) \approx P(\bar{X} > 16.024922).$$

We can solve the above to find the value of m where the above is true. We have

$$\frac{1}{4} = P(\bar{X} > 16.024922) = 1 - P(\bar{X} < 16.024922) \quad \text{so} \quad P(\bar{X} < 16.024922) = \frac{3}{4}.$$

This is equivalent to

$$P\left(\frac{\bar{X} - m}{3/\sqrt{5}} < \frac{16.024922 - m}{3/\sqrt{5}}\right) = \frac{3}{4},$$

or

$$\Phi\left(\frac{16.024922 - m}{3/\sqrt{5}}\right) = \frac{3}{4}.$$

Solving for m we find $m = 15.119999$.

Part (c): In this case we would have

$$\bar{X} \sim \mathcal{N}\left(14, \frac{\sigma^2}{5}\right),$$

with σ unknown. The new control limits with this new σ are now given by

$$12 \pm 3\frac{\sigma}{\sqrt{5}}.$$

Following similar steps as in the above (assuming the left tail has negligible probability) we would get the equation

$$\Phi\left(\frac{12 + 3\frac{\sigma}{\sqrt{5}} - 14}{\sigma/\sqrt{5}}\right) = \frac{3}{4}.$$

Solving this for σ we find $\sigma = 1.923077$. We would next need to verify that the left tail has a negligible probability and the above approximations are good (not shown).

Part (d): The upper control limit would become $\text{UCL} = 12 + 3\left(\frac{3}{\sqrt{n}}\right)$. Following the results above we would have

$$\Phi\left(\frac{\text{UCL} - 14}{3/\sqrt{n}}\right) = \frac{3}{4}.$$

Solving this for n we find $n = 12.167995$ which we would round upwards to the integer 13.

Exercise 6

Part (a): The upper and lower control limits are given by

$$\mu \pm 3\frac{\sigma}{\sqrt{n}}.$$

For the numbers given in this problem these are

```
[1] "LCL= 5.000000; UCL= 11.000000"
```

If the process mean shifts to 9 then we will now have our sample means $\bar{X}$ distributed as

$$\bar{X} \sim \mathcal{N}\left(9, \frac{\sigma^2}{n}\right).$$

To determine the average number of samples to be drawn until the shift in the mean is observed one needs to compute the probability that a single sample (of size n) exceeds the control limits above or

$$p = P(\bar{X} \leq \text{LCL}) + P(\bar{X} \geq \text{UCL}).$$

As the shift is upwards we will approximate p as $p \approx P(\bar{X} \geq \text{UCL})$. We can write this as

$$P(\bar{X} \geq \text{UCL}) = 1 - P(\bar{X} < \text{UCL}) = 1 - P\left(\frac{\bar{X} - 9}{\sigma/\sqrt{n}} < \frac{\text{UCL} - 9}{\sigma/\sqrt{n}}\right),$$

and we can evaluate the right-hand-side using standard normal theory. Doing this we find $p = 0.022750$. The average number of samples we need to observe before we see a control limit breached is $\frac{1}{p} = 43.955789$.

Part (b): Let m be the upward shift. Then $\bar{X} \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$ and we desire to find m such that

$$p = \frac{1}{8} = P(\bar{X} \geq \text{UCL}).$$

Transforming to a standard normal random variable we would like to find m such that

$$P\left(\frac{\bar{X} - m}{\sigma/\sqrt{n}} < \frac{\text{UCL} - m}{\sigma/\sqrt{n}}\right) = \frac{7}{8}.$$

Solving the above for m we find $m = 9.849651$.

Part (c): Let the new standard deviation be σ. Then the upper control limit is given by

$$\text{UCL} = 8 + \frac{3\sigma}{\sqrt{n}}.$$

To have an ARL of 8 and expecting an upwards shift in the mean (i.e. we expect our upper control limit to be breached) means that

$$P(\bar{X} > \text{UCL}) \approx \frac{1}{8},$$

or

$$P(\bar{X} < \text{UCL}) \approx \frac{7}{8}.$$

As we know that under the proposed shift in the mean we have that $\bar{X} \sim \mathcal{N}\left(9, \frac{\sigma^2}{n}\right)$ we can use normal theory to evaluate the above and solve for σ. Doing this we find $\sigma = 1.081285$.

Part (d): Let the new standard samples size be n. Then the upper control limit is given by

$$\text{UCL} = 8 + 3\frac{2}{\sqrt{n}}.$$

By using the logic in the previous part of this exercise to have an ARL of 8 and expecting an upwards shift in the mean we would desire

$$P(\bar{X} < \text{UCL}) \approx \frac{7}{8}.$$

As we know that under the proposed shift in the mean we have that $\bar{X} \sim \mathcal{N}\left(9, \frac{4}{n}\right)$ we can use normal theory to evaluate the above and solve for n. Doing this we find $n = 13.684830$ which we would round upwards to the integer 14.

Exercise 7

Part (a-c): The probability that a single sample is outside of the control limits is

$$p = P(\bar{X} \leq \text{LCL}) + P(\bar{X} \leq \text{UCL}).$$

As $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ we have

$$\text{LCL} = \mu - \frac{3\sigma}{\sqrt{n}}$$
$$\text{UCL} = \mu + \frac{3\sigma}{\sqrt{n}}.$$

This means that

$$P(\bar{X} \leq \text{LCL}) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq -3\right) = \Phi(-3) = 0.001350,$$

and by symmetry p will be twice this value. The probability that a false alarm will occur at least once in the next N samples is the complement to the probability that a false alarm will *not* occur in the next N samples. This is

$$1 - (1 - p)^N.$$

For the Ns asked about in this problem we have the probabilities given by

```
[1] 0.1264347 0.2368836 0.4176534
```

Part (d): For this we would want to find N such that

$$1 - (1 - p)^N = 0.5.$$

We find $N = 256.393840$ which we would round to the integer 257.

Exercise 8

This problem is worked in the R code `chap_10_sect_2_prob_8_N_9.R`.

Part (a): Plotting the R chart we don't see any out of control samples.

Part (b): Plotting the $\bar{X}$ chart based on the sample range we see that the third sample is above the 3σ limits and the process can be declared out of control.

Part (c): Using the the $\bar{X}$ chart based on the sample range we see that two samples are beyond the 3σ limits (numbers three and ten). In addition, samples 4 and 5 are below the -2σ limit triggering a two-out-of-three violation.

Exercise 9

Part (a): Plotting the S chart we don't see any out of control samples.

Part (b): Plotting the $\bar{X}$ chart based on the sample standard deviation we see that the third sample is above the 3σ limits and the process can be declared out of control.

Part (c): Using the the $\bar{X}$ chart based on the sample standard deviation we see that two samples are beyond the 3σ limits (numbers three and ten). In addition samples 4 and 5 are below the -2σ limit triggering a two-out-of-three violation. These are the same conclusions reached by plotting the $\bar{X}$ control chart but based on the sample range.

Exercise 10

This problem is worked in the R code `chap_10_sect_2_prob_10_N_11.R`.

Part (a): Plotting the R chart we don't see any out of control samples.

Part (b): Plotting the $\bar{X}$ chart based on the sample range we see that no sample is outside the 3σ limits and the process can be declared "in control".

Part (c): Using the the $\bar{X}$ chart based on the sample range we see that samples 7 and 9 are beyond the -2σ limits triggering a two-out-of-three violation. Starting at sample 8 we will have eight consecutive points on the same side (below) the center line.

Exercise 11

Part (a): Plotting the S chart we don't see any out of control samples.

Part (b): Plotting the $\bar{X}$ chart based on the sample standard deviation we see that no sample is outside the 3σ limits and the process can be declared "in control".

Part (c): Using the the $\bar{X}$ chart based on the sample standard deviation we see that samples 7 and 9 are beyond the -2σ limits triggering a two-out-of-three violation. Starting at sample 8 we also have eight consecutive points on the same side (below) the center line.

Exercise 12

This problem is worked in the R code `chap_10_sect_2_prob_12_N_13.R`.

Part (a): Plotting the R chart we see that sample 8 has a range larger than the 3σ limits and can be considered out of control. This sample will need to be removed. We find

```
[1] "W/O sample 8: X_barbar= 150.165517; R_bar= 6.537931; s_bar= 2.911379"
```

Part (b): Plotting the $\bar{X}$ chart based on the sample range we see that no sample is outside the 3σ limits and the process can be declared "in control".

Part (c): Using the the $\bar{X}$ chart based on the sample range we see that the Western Electric rules are not violated and again the process can be declared "in control".

Exercise 13

Part (a): Plotting the S chart we see that sample 8 has a range larger than the 3σ limits and can be considered out of control. This sample will need to be removed as in the previous Exercise.

Part (b): Plotting the $\bar{X}$ chart based on the sample standard deviation we see that no sample is outside the 3σ limits and the process can be declared "in control".

Part (c): Using the the $\bar{X}$ chart based on the sample range we see that the Western Electric rules are not violated and again the process can be declared "in control".

Control Charts for Attributes

Exercise 1

The center line would be centered on

$$\bar{p} = \frac{1}{30} \sum_{i=1}^{30} p_i = 0.05466667,$$

with

$$\sigma = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} = 0.04150431.$$

This means that the lower, center, and upper line of the p chart are located at

```
[1] 0.006443021 0.054666667 0.102890312
```

These are calculated in the following R code

```
p_bar = 1.64/30
n = 200
sigma = sqrt(p_bar*(1-p_bar)/n)
c(p_bar - 3*sigma, p_bar, p_bar + 3*sigma)
```

Exercise 2

Part (a): In the `R` code for this exercise we find

```
[1] 0.04003114 0.08946667 0.13890220
```

Part (b): Running the `R` code for this exercise plots the sample proportion as a function of the sample number. From that plot one can see that the second and the 24th sample are beyond the lower or upper control limits indicating that the process is not in control at that location. The sample proportion seems to increase with the sample index. The first time this sample can be declared "out-of-control" is on the second sample.

Exercise 3

Our sample size has $n = 100$. In each sample, the sample proportion is the number X_i of defective items divided by n so

$$\hat{p}_i = \frac{X_i}{n}\,.$$

We can estimate p (the true defective proportion) by averaging $k = 50$ sample values of $\hat{p}_i$. We have

$$\bar{p} = \frac{1}{k}\sum_{i=1}^{k}\frac{X_i}{n} = \frac{1}{nk}\sum_{i=l}^{k}X_i = \frac{1}{nk}(622) = \frac{622}{5000} = 0.1244\,.$$

The estimate of σ is given by

$$\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.03300373\,.$$

Thus we do have enough information to compute a p control chart. I find that the lower, center, and upper line of the p chart are located at

```
[1] 0.0253888 0.1244000 0.2234112
```

These calculations are done in the following `R` code

```
n = 100
k = 50
p_bar = 622/(n*k)
sigma = sqrt(p_bar*(1-p_bar) /n)
c(p_bar - 3*sigma, p_bar, p_bar + 3* sigma)
```

Exercise 4

The two given samples have proportions given by $\hat{p} = \frac{24}{100} = 0.24$ and $\hat{p} = \frac{6}{100} = 0.06$. As the larger one is beyond the upper confidence bound (presented above) we must conclude that this process is not in control.

Exercise 5

Here our samples size is $n = 20$ and an estimate of the defect probability p is given by

$$\hat{p} = \frac{1}{k} \left(\sum_{i=1}^{k} \frac{X_i}{n} \right) = \frac{1}{10} \left(\sum_{i=1}^{10} \frac{X_i}{20} \right) = \frac{1}{200}(3) = 0.015 \,.$$

As $np = 0.3 < 10$ our control limits will not be valid. The easiest way to correct this would be to increase our sample size n.

Exercise 6

We first compute

$$\bar{c} = \frac{1}{k} \sum_{i=1}^{k} c_i = \frac{876}{70} = 12.5143 \,.$$

Then the lower, center, and upper control limits are given by $\bar{c} - 3\sqrt{\bar{c}}$, $\bar{c}$, and $\bar{c} + 3\sqrt{\bar{c}}$. These are the numbers

```
[1]  1.901625 12.514286 23.126947
```

Computed with the simple R code

```
c_bar = 876/70
c_bar + c(-3*sqrt(c_bar), 0, +3*sqrt(c_bar))
```

Exercise 7

Notice that 34 is larger than the upper control limit computed in the previous exercise. Thus we can conclude that this process is *not* in control.

Exercise 8

Part (a): Using the R code for this exercise we compute the needed figures. Running that code I find

```
[1] "c_bar= 45.300000; sigma= 6.730527"
[1] 25.10842 45.30000 65.49158
```

Part (b): Running the R code for this gives a c chart for this data. All points seem to be inside the lower and upper confidence interval limits indicating that the process is in control.

The CUSUM Chart

Exercise 1

This problem is worked in the R code `chap_10_sect_4_prob_1.R`.

Part (a): No samples need to be deleted to bring the process variation under control.

Part (b): I compute $\sigma_{\bar{X}} = 0.027$.

Part (c): Running the above R code gives a CUSUM plot.

Part (d): The process mean will be declared out-of-control if either SH_i or SL_i is larger than $h\sigma_{\bar{X}}$ (in absolute value). From the plot above we see that the *seventh* sample is "almost" beyond $-h\sigma_{\bar{X}}$ while the *eighth* sample is beyond this level (the first sample in the plot is at index zero).

Part (e): This was done in the previous problem where we found that at sample seven the process is out-of-control.

Exercise 2

This problem is worked in the R code `chap_10_sect_4_prob_2.R`.

Part (a): No samples need to be deleted to bring the process variation under control.

Part (b): I compute $\sigma_{\bar{X}} = 0.033$.

Part (c): Running the above R code gives a CUSUM plot.

Part (d): The process mean will be declared out-of-control if either SH_i or SL_i is larger than $h\sigma_{\bar{X}}$ (in absolute value). From the plot above we see that the *fifth* sample is "almost" beyond $-h\sigma_{\bar{X}}$ while the *tenth* sample is beyond this level (the first sample in the plot is at index zero).

Part (e): This was done in the previous problem where the $\bar{X}$ chart there indicated that the process was out-of-control at the third sample which was an earlier detection than the CUSUM detection here.

Exercise 3

This problem is worked in the R code `chap_10_sect_4_prob_3.R`.

Part (a): No samples need to be deleted to bring the process variation under control.

Part (b): I compute $\sigma_{\bar{X}} = 0.219$.

Part (c): Running the above R code gives a CUSUM plot.

Part (d): The process mean will be declared out-of-control if either SH_i or SL_i is larger than $h\sigma_{\bar{X}}$ (in absolute value). From the plot above we see that the *ninth* sample is "almost" beyond $-h\sigma_{\bar{X}}$ while the *tenth* sample is beyond this level (the first sample in the plot is at index zero).

Part (e): This was done in the previous problem where the $\bar{X}$ chart there indicated that the process was out-of-control at the ninth sample (where two of three samples were beyond the 2σ limits). That is slightly earlier than the CUSUM detection.

Exercise 4

This problem is worked in the R code `chap_10_sect_4_prob_4.R`.

Part (a): The eighth sample needed to be dropped to bring the process variation under control.

Part (b): I compute $\sigma_{\bar{X}} = 1.589$.

Part (c): Running the above R code gives a CUSUM plot.

Part (d): The process mean will be declared out-of-control if either SH_i or SL_i is larger than $h\sigma_{\bar{X}}$ (in absolute value). From the plot above we see that none of the values of SH_i or SL_i is that large. Thus we conclude that the process mean is in control.

Part (e): The Western Electric rules found the process to be in control also.

Exercise 5

This problem is worked in the R code chap_10_sect_4_prob_5.R.

Part (a): Running the above R code gives a CUSUM plot.

Part (b): The process mean will be declared out-of-control if either SH_i or SL_i is larger than $h\sigma_{\bar{X}}$ (in absolute value). From the plot above we see that none of the values of SH_i or SL_i is that large. Thus we conclude that the process mean is in control.

Process Capability

Exercise 1

From the formulas in the book we have

$$C_{pk} = \min\left(\frac{\hat{\mu} - \text{LSL}}{3\hat{\sigma}}, \frac{\text{USL} - \hat{\mu}}{3\hat{\sigma}} \right).$$

For the numbers given in this problem we first estimate $\hat{\sigma} = \frac{\bar{s}}{c_4} = \frac{0.002}{0.9515} = 0.002101944$. Then the two numbers above are

$$\frac{0.248 - 0.246}{3(0.002101944)} = 0.3171667$$

$$\frac{0.254 - 0.248}{3(0.002101944)} = 0.9515000.$$

The smaller is the first so $C_{pk} = 0.3171667$. As $C_{pk} < 1$ the process capability is not acceptable.

Exercise 2

From the formulas in the book we have

$$C_{pk} = \min\left(\frac{\hat{\mu} - \text{LSL}}{3\hat{\sigma}}, \frac{\text{USL} - \hat{\mu}}{3\hat{\sigma}} \right).$$

For the numbers given in this problem we first estimate $\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{0.05}{2.847} = 0.01756235$. Then the two numbers above are

$$\frac{15.52 - 15.40}{3(0.01756235)} = 2.2776$$

$$\frac{15.60 - 15.51}{3(0.01756235)} = 1.5184 \,.$$

The smaller is the second so $C_{pk} = 1.5184$. As $C_{pk} > 1$ the process capability is acceptable.

Exercise 3

To maximize the process capacity the process mean should be set to the midpoint between the lower specification limit and the upper specification limit i.e. at

$$\frac{\text{LSL} + \text{USL}}{2} \,.$$

For the data here this is the value 15.5. The process capability will then be $C_p = \frac{\text{USL} - \text{LSL}}{6\hat{\sigma}} = 1.898$.

Exercise 4

Part (a): To maximize the process capacity the process mean should be set to the midpoint between the lower specification limit and the upper specification limit i.e. at

$$\frac{\text{LSL} + \text{USL}}{2} \,.$$

For the data here this is the value 0.25.

Part (b): With that change the process capability will then be $C_p = \frac{\text{USL} - \text{LSL}}{6\hat{\sigma}} = 0.6343333$. As $C_p < 1$ the process cannot be made acceptable by only adjusting the process mean.

Part (c): We would need to have

$$\frac{\text{USL} - \text{LSL}}{6\hat{\sigma}} \geq 1 \quad \text{so} \quad \hat{\sigma} \leq \frac{\text{USL} - \text{LSL}}{6} = 0.001333333 \,.$$

Part (d): To be "six-sigma quality" we must have

$$C_{pk} \geq 2 \quad \text{or} \quad \frac{\text{USL} - \text{LSL}}{6\hat{\sigma}} \geq 2 \quad \text{or} \quad \hat{\sigma} \leq \frac{\text{USL} - \text{LSL}}{12} = 0.0006666667 \,.$$

Exercise 5

Part (a): We are told that
$$\hat{\mu} = \frac{\text{LSL} + \text{USL}}{2} \,,$$
and that because of this that
$$C_p = \frac{\text{USL} - \text{LSL}}{6\hat{\sigma}} = 1.2 \,.$$
From these we get
$$\text{USL} + \text{LSL} = 2\hat{\mu}$$
$$\text{USL} - \text{LSL} = 6C_p\hat{\sigma} \,.$$
Solving we find
$$\text{USL} = \hat{\mu} + 3C_p\hat{\sigma}$$
$$\text{LSL} = \hat{\mu} - 3C_p\hat{\sigma} \,.$$

Part (b): We want to estimate p where
$$p = P((X_i < \text{LSL}) \cup (X_i > \text{USL})) = 2P(X_i < \text{LSL}) \quad \text{by symmetry}$$
$$= 2P\left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} < \frac{\text{LSL} - \hat{\mu}}{\hat{\sigma}}\right) \,.$$
Now when LSL is given by the above expression we find
$$p = 2P(Z < -3C_p) = 2\Phi(-3C_p) \,.$$
Evaluating this in `R` using

```
C_p = 1.2
2*pnorm(-3*C_p)
```

give $p = 0.0003182172$ which is slighly different than the answer given in the back of the book. If anyone sees anything in error please let me know.

Part (c): It is likely that the true proportion of nonconforming units will be different than this number because the normal approximation is expected to be "poor" in the tails of the distribution.

Supplementary Exercise

Exercise 1

We can do this with the following `R` code

```
p_bar = 5.83/100
n = 300
sigma_hat = sqrt(p_bar*(1-p_bar)/n)
print(c(p_bar - 3*sigma_hat, p_bar, p_bar + 3*sigma_hat))
```

This gives

```
[1] 0.01771634 0.05830000 0.09888366
```

Exercise 2

Part (a): If p is the probability that a sample mean $\bar{X}$ is outside of the control limits then we have

$$p = P((\bar{X} < \mu - 2\sigma_{\bar{X}}) \cup (\bar{X} > \mu + 2\sigma_{\bar{X}})) = 2P(\bar{X} < \mu - 2\sigma_{\bar{X}}) \quad \text{by symmetry}$$
$$= 2P\left(\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} < -2\right) = 2\Phi(-2).$$

Evaluating this in R using

```
2*pnorm(-2)
```

give $p = 0.04550026$. Thus ARL $= \frac{1}{p} = 21.97789$.

Part (b): The new process mean will be $\mu' = \mu - 0.5\sigma_{\bar{X}}$ (so $\mu - \mu' = 0.5\sigma_{\bar{X}}$) and so using μ' the above calculation is

$$p = P((\bar{X} < \mu - 2\sigma_{\bar{X}}) \cup (\bar{X} > \mu + 2\sigma_{\bar{X}})) = P(\bar{X} < \mu - 2\sigma_{\bar{X}}) + P(\bar{X} > \mu + 2\sigma_{\bar{X}})$$
$$= P\left(\frac{\bar{X} - \mu'}{\sigma_{\bar{X}}} < \frac{\mu - \mu' - 2\sigma_{\bar{X}}}{\sigma_{\bar{X}}}\right) + P\left(\frac{\bar{X} - \mu'}{\sigma_{\bar{X}}} > \frac{\mu - \mu' + 2\sigma_{\bar{X}}}{\sigma_{\bar{X}}}\right).$$

Simplifying this we get

$$p = P\left(Z < -\frac{3}{2}\right) + P\left(Z > \frac{5}{2}\right) = \Phi\left(-\frac{3}{2}\right) + \Phi\left(-\frac{5}{2}\right) = 0.07301687,$$

using

```
pnorm(-3/2) + pnorm(-5/2)
```

to evaluate the cumulative standard normal. This means that ARL $= \frac{1}{p} = 13.69547$.

Part (c):

Exercise 3

Exercise 4

Exercise 5

Exercise 6

Exercise 7